

PB83-247957

Testing Models of the Fate of
Chemicals in Aquatic Environments

(U.S.) Environmental Research Lab., Athens, GA

1983



U.S. Department of Commerce
National Technical Information Service

NTIS

TECHNICAL REPORT DATA		
(Please read instructions on the reverse before completing)		
1. REPORT NO. EPA-600/D-83-100	2.	3. RECIPIENT'S ACCESSION NO. PB8 3 247957
4. TITLE AND SUBTITLE Testing Models of the Fate of Chemicals in Aquatic Environments	5. REPORT DATE 1983	
	6. PERFORMING ORGANIZATION CODE	
7. AUTHOR(S) Ray R. Lassiter	8. PERFORMING ORGANIZATION REPORT NO.	
9. PERFORMING ORGANIZATION NAME AND ADDRESS Environmental Research Laboratory U.S. Environmental Protection Agency College Station Road Athens GA 30613	10. PROGRAM ELEMENT NO. ACBELA	
	11. CONTRACT/GRANT NO.	
12. SPONSORING AGENCY NAME AND ADDRESS U.S. Environmental Protection Agency--Athens GA Office of Research and Development Environmental Research Laboratory Athens GA 30613	13. TYPE OF REPORT AND PERIOD COVERED	
	14. SPONSORING AGENCY CODE EPA/600/01	
15. SUPPLEMENTARY NOTES		
16. ABSTRACT Laboratory ecosystems have been used to test a model for the fate of toxic chemicals. Two questions are asked in such a test. First, are the laboratory systems functional ecosystems in which primary productivity, secondary productivity, decomposition, element cycling, and various synergisms occur? Second, is the model capable of predicting the fate of the chemical in the laboratory ecosystem? The answer to the first question is important in the interpretation and potential extrapolation of the answer to the second. These relationships have been considered and evaluated with the conclusion that laboratory systems can be ecosystems provided certain operational requirements are met. The model (the basis of the Exposure Analysis Modeling System, EXAMS) is not rejected using ordinary statistical analysis, and, based on arguments founded upon fundamental concepts in physical chemistry, it is accepted as a working theory of the behavior of chemicals in the environment.		
17. KEY WORDS AND DOCUMENT ANALYSIS		
a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group
19. DISTRIBUTION STATEMENT RELEASE TO PUBLIC	19. SECURITY CLASS (This Report) UNCLASSIFIED	21. NO. OF PAGES 22
	20. SECURITY CLASS (This page) UNCLASSIFIED	22. PRICE

Testing Models of the Fate of Chemicals in Aquatic Environments

by

Ray R. Lassiter *
Institute of Ecology
University of Georgia
Athens, Georgia 30602

Paper presented at the workshop "Modeling the Fate of Chemicals in the Aquatic Environment", August 17 - 21, 1981, Pellston, Michigan.

* Address correspondence to: Environmental Research Laboratory
Office of Research and Development
U.S. Environmental Protection Agency
Athens, GA 30613

REPRODUCED BY
NATIONAL TECHNICAL
INFORMATION SERVICE
U.S. DEPARTMENT OF COMMERCE
SPRINGFIELD, VA. 22161

ENVIRONMENTAL RESEARCH LABORATORY
OFFICE OF RESEARCH AND DEVELOPMENT
U.S. ENVIRONMENTAL PROTECTION AGENCY
ATHENS, GA 30613

NOTICE

This document has been reviewed in accordance with U.S. Environmental Protection Agency policy and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

ABSTRACT.

Assessment of the fate of chemicals in the aquatic environment is largely concerned with new chemicals. For these and other chemicals whose environmental fate is unknown, predictions can only be based upon theory. Theoretical models are used in these assessments. Tests of these models necessarily are carried out at three levels: the underlying theory, the faithfulness with which the model represents the theory, and the utility of the model. For models of the fate of chemicals in aquatic environments, much underlying theory in relevant disciplines has already been developed and tested. Tests for model faithfulness to theory (concept) and for utility recently have been carried out. The Exposure Analysis Modeling System (EXAMS) was tested for concept and utility using methyl parathion in laboratory ecosystems. To test the concept, the model was parameterized with system-specific parameters. Differences between predicted and observed concentrations of methyl parathion, then, would necessarily be interpreted to have been derived from misrepresentation or omission of theory. For utility, tests were carried out using averaged parameters, such as might be necessary in practice. It was concluded that EXAMS is conceptually sound. Based upon the finding that biological decomposition varies widely in environments dominated by aufwuchs, it was suggested that EXAMS utility is reduced for assessing certain chemicals in those environments. Models must finally be tested for their adequacy for their designed purpose. The end user must be the final judge of adequacy.

INTRODUCTION

It is generally accepted that models ought to be tested before they are put to serious application. The way that models ought to be tested, however, is not widely agreed on, nor even frequently discussed. I shall present the view of models and model testing that underlay and motivated our testing the Exposure Analysis Modeling System (EXAMS). In predicting the fate of chemicals in the aquatic environment, both empirical and theoretical models have been used. Empirical models are discussed by others in this conference, and I shall consider them only to aid in discussing theoretical models. Both classes of models, in the final analysis, are tools that, in the present context, are used to predict some aspect of a chemical's fate in the aquatic environment. In the past both empirical and theoretical models have been needed in this field, and there is reason to believe that it will continue to be so well into the future.

Different kinds of models and different uses of models need to be tested differently. One needs to test whether a model is adequate for its purpose. As a basis for considering what aspects need to be tested for models of the environmental fate of chemicals, I shall present, briefly, my perception of the way that theoretical models relate to theory, and what is necessary for them to be applied in practice. A brief and very general description of empirical models, as applied in environmental science, should help to clarify the differences by contrast between empirical and theoretical models.

EMPIRICAL MODELS

Caswell (1976) refers to empirical models as predictive models. These models are entirely utilitarian; their sole purpose being to produce useful predictions. Therefore one cannot validly test or even question their theoretical basis. Only the adequacy of the results can be criticized. A regression model is one of the simplest cases. In environmental science the various calibrated models are empirical and predictive. Calibration (fitting) involves varying the parameters of the model until the differences between the predictions and observations satisfy some criterion, e. g., the minimum sum of squares. Such models might employ terms intended to reflect underlying theory, but if they depend upon calibration for their predictive power, then they are empirical. Testing these models entails establishing their adequacy to satisfy the purpose for which they were created. Whether such models are adequate in a given instance depends heavily upon the statistical characteristics of the calibration data set. The description of empirical models as predictive has to be understood in a special way. There is no intent for such a model to be used to make predictions outside the range for which it was calibrated. Indeed there is no basis for credibility of such a prediction. Therefore, an empirical model's prediction is in some sense more akin to an interpolation. Of course there may be many specific

predictions within the range of calibration that have never been observed and/or which could not be deduced without the model. In fact the utility of empirical models lies in its ability to make such predictions.

THEORETICAL MODELS

Theoretical models are distinguished from empirical models by the way in which their parameters receive values. A parameter value of a theoretical model is associated with a term that represents a process. Its value is estimated in some physical reference to that process, that is, in an experiment or other task in which measurements are made on relevant aspects of the process. Never is a contrast made between predictions by the model and observations of what is predicted for the purpose of establishing a value of any parameter. Therefore, predictions of a theoretical model are not limited by the range of observations of a data set. They are limited by the generality of the terms that represent the processes and by uncertainty associated with estimates of the process parameters. Of course, they also are potentially limited by mistakes and omissions made when the model was constructed. Apart from these potential limitations, no bounds are placed upon the range of predictions that a theoretical model can legitimately make. It is, therefore, these limitations that need to be examined when theoretical models are tested.

Theoretical models can be utilitarian, or, as Caswell (1976) puts it, their purpose can be "to gain insight (*italics*) into how the system operates." I prefer to distinguish between model and theory, and would modify his statement to read, "to gain insight into how theory operates," or perhaps "... into how theory implies that the system operates." Theoretical models are supposed to be conceptually correct by their presumed correspondence to underlying theory. A theory is a body of concepts and relationships explaining some aspect of the world. These concepts and relationships can be used to make observable predictions about aspects of the world that are not explicit in the theory. Some of the concepts and relationships of a particular theory originate in other theories. We generally use "theory" to refer generically to the body of concepts, relationships, and implications associated with theories. Models, themselves, are not theories, however. A specific, theoretical model is one of many possible instruments by which theory can be translated into specific cases (predictions). A single equation or logical relationship can be considered a model if it is used to obtain a result implied by theory. At the other end of the spectrum is the large set of coupled or otherwise interconnected equations and logic too complicated for paper and pencil manipulation and, therefore, manipulated by computer. Currently, particularly in environmental science and related fields, the latter is called a model, whereas a single equation or logical relationship often is not. Every such expression of theory, however, legitimately can be called a theoretical model.

Overton's (1977) view of the modeling process is that it consists of searching for appropriate model components to accomplish two objectives. First the model predictions must match observed values ("prescribed behavior"). Second it must incorporate sufficient theory to permit extrapolation. Thus models can interact with theories in two directions. That is, they can be based upon theory, but ultimately theory, also, can be based upon a model, though in an indirect way. A theory is not necessarily complete; yet a model, if it is to make concrete predictions, must be logically complete. Where the basis for a model does not exist in underlying theory, the modeler is obliged to improvise. The improvisation can take the form of an empirical submodel, or, if warranted, the modeler can complete the model by using a conjectural, theoretical construct to substitute for the nonexistent theory. In keeping with Popper's (1959, 1962) theory of the growth of scientific knowledge, such a conjecture, if submitted to repeated experiments designed to refute it should it in fact be false, can become accepted as part of the underlying theory. In environmental science the underlying theory consists of relevant portions of the theories of physics, chemistry, biology, geology, meteorology, and perhaps many other disciplines.

TESTING THEORETICAL MODELS

The essence of testing a model is a comparison of its predictions to observations of the phenomena that it predicted, followed by analysis, conclusions about the model, and, if weaknesses are detected, conjectures for revising the model to overcome the weaknesses. The intent of testing, as already alluded to, is to subject the model to the severest tests conceivable, so as to expose any deficiencies or fallacies as soon as possible. Of equal importance to exposing errors is the subsequent revision of the model with the intent that the revised model too will be subjected to equally critical tests. Popper (1962) refers to it as essentially a method of trial and error, or more explicitly, "conjecture and refutation."

The method of science, however, differs from simple trial and error by conscious efforts to uncover errors in the testing stage so as to avoid their turning up inopportunistically in the use stage. Popper has effectively laid to rest any notion that a theory (and therefore the basis for a theoretical model) can be validated in a positive sense. One can show that a model's predictions were accurate in a given instance, or several instances, but one can never show that a model or its underlying theory is correct, in general. At best the credibility or degree of belief that it will predict accurately for all future cases can be improved. As Polya (1953) puts it, verification in a singular case makes the general case "more plausible."

THE EXPOSURE ANALYSIS MODELING SYSTEM

We constructed the Exposure Analysis Modeling System (EXAMS) to make evaluative analyses of the fate of organic chemicals in

aquatic environments (Burns et al., 1981). To make an evaluative analysis of a chemical to be released to the environment, we visualized calculating its expected fate using environment representations that span the characteristics of the real environments that the chemical is expected to enter. These environment representations, of course, do not correspond to specific sites. Analyses of this sort can be useful only if calculations are accurate for an environment as described. That is, if an environment did exist exactly as described in the evaluative analysis, then predicted concentrations should match the observed ones. Otherwise predictions would be misleading and therefore of negative value to one charged with making a regulatory decision. In general the environments used in evaluative analyses are fictitious, and it is a hypothetical question to ask whether predictions would match observations on the system. However, some existing environments, notably laboratory ecosystems, can be characterized in the manner required by EXAMS, and therefore, they can be used to provide observations against which to compare the model's predictions. Accordingly, we used laboratory ecosystems maintained with several different environments to make the tests.

Our first concern was to test whether EXAMS was theoretically sound. If the model perfectly described an environment and the transport and transformations of a chemical in it, predictions for fate of that chemical would perfectly match the chemical's actual fate. In testing EXAMS for its conceptual validity, we used this idea as a pattern. We hypothesized that the model's structure and theoretical content was correct and supplied the model with the best available parameters. The best available parameters were those that corresponded most closely to the sites and times where we made the observations on chemical fate. Similarly, we used specific chemistry parameters for the test chemical, methyl parathion. For biological decomposition, a separate second order rate coefficient was estimated and used for each of the experimental environments. By this means the parameterized model was, as nearly as possible, a specific representation of the test systems. Thus any differences between model and observation should be attributable to fallacies in the model's structure or theoretical content. We formulated the null hypothesis to reflect these ideas: there is no difference between predicted and observed concentrations when the model is parameterized with environment-specific parameters. Results of this test were analyzed by linear regression of observed on predicted concentrations. The null hypothesis implies a linear relationship with intercept of zero and slope of one. Figure 1 is a display of the data plotted along with this hypothetical line. Student's *t* was used to test whether slope and intercept differ from their hypothesized values. These results are given in Table 1. No differences can be detected, and therefore the null hypotheses are not rejected. As a result of this experimental test, the model has increased in plausibility.

In outline, this approach to testing consists of the following

steps:

- ... design the content of the model test,
- ... devise a statement of no difference between model predictions and observations,
- ... use regression of observations on predictions and test working level null hypotheses of intercept = 0 and slope = 1,
- ... interpret tests and draw conclusions.

Designing the content of the model is the step requiring the greatest comprehension of the problem at hand. In testing EIXMS we established content of the test by the assumptions that we used to make the predictions. In the test described above, we used the most specific parameters that we had available, under the assumption that the model, if conceptually sound, would thereby provide the most accurate predictions. In another test we used the same experiment with different assumptions (more observations qualified) to test the model's utility.

Some aspects of model utility will need to be checked with reference to the specific application. In general, though, no user will have access to site- and situation-specific parameters with which to run the model. To the extent that situation-specific parameters are necessary to provide useful results in practice, the utility of the model will be limited. This suggests that one useful test of utility is to remove any situation-specific parameters from the model and test the hypothesis of no difference using situation-independent parameterization. We know of no evidence to suggest that abiotic chemistry processes vary as a function of environment. That is, these rate coefficients and equilibrium constants are considered to be situation-independent. It is frequently suggested, however, that biological decomposition rate coefficients vary with environment, and accordingly, in the test for concept we used situation-specific values for these rate coefficients, as already discussed. To test utility, however, we obtained a rate coefficient applying to all environments by averaging the situation specific rate coefficients over environments. The null hypothesis was -- there is no difference between predicted and observed concentrations when a single biological decomposition rate coefficient is used for all environments.

Regression analysis was used as in the outline above. Therefore, the null hypothesis was translated into working null hypotheses, as in the test for concept, that intercept = 0 and slope = 1. A plot of observed versus predicted concentrations is shown in Figure 2, and results of the test are presented in Table 2. There is little doubt that the null hypotheses should be rejected and the conclusion drawn that the predictions are detectably different from the observations.

The model has been accepted as conceptually sound, but rejected for use with environmentally invariant biological decomposition

rate coefficients. To interpret this rejection in terms of model utility, however, additional factors are relevant. What is desired as a result of the testing is to be able to establish bounds upon the situations for which EXAMS is applicable. We can say from this single test that the biological decomposition rate coefficient is not universally invariant with respect to environment. On the other hand it is not prudent, and certainly not logically justifiable, to extend the interpretation much beyond that. Our test system was heavily dominated by aufwuchs, and we have not tested the invariance of biological decomposition rate coefficients in other classes of systems. There is also evidence that test results were influenced by major shifts in the biological community prior to the last sampling period. Paris et al. (1981) have studied second order biological decomposition rates in planktonic systems. They studied three chemicals in more than 40 water bodies. Overall they found a coefficient of variation of about 65%. It is difficult to say just how much variation in model predictions would result from rate coefficients varying over a range equivalent to a coefficient of variation of 65%. It will probably be substantially less than the variation noted using rate coefficients from aufwuchs communities as in our tests, however.

Two recommendations are appropriate based upon the above considerations. First, EXAMS needs to be tested more along these same lines, using different chemicals and environments. Second, when one uses EXAMS to make analyses for systems dominated by aufwuchs, he should exercise caution. Of particular concern are chemicals that are slow to degrade and whose major pathway of degradation is microbial decomposition. In any event one can exercise the simple expedient of doing two or more analyses to obtain a range of predicted values when using parameters with associated uncertainty.

The potential limitation of EXAMS' utility for analyzing the fate of chemicals in systems dominated by aufwuchs naturally led to questions of how to overcome the limitation. Such questions are the beginnings of the process that can eventually lead to conjectures for revising the model. Popper (1959) argues that new hypotheses cannot be originated mechanically or analytically, but are purely conjectural. It is reasonable, however, to pursue analyses with the intent and expectation of stimulating ideas. A serious conjecture for revising the model can have its beginning in just such an idea. Differences between observations and predictions conceivably can be associated with factors of the test experiment in ways that are suggestive of model improvement. For EXAMS we analyzed the differences using the predictions made under the assumption of invariance of the biological decomposition parameter. Table 3 contains the analysis of variance for these data.

Several interesting avenues of discussion are possible based on the results presented in Table 3. These are discussed elsewhere (Lassiter et al., in prep.), but the clear association of CSTR (Continuously Stirred Tank Reactors, maintained as different environments) with differences is of particular interest here.

We examined this association more closely using Duncan's Multiple Range test (Table 4). As a result of this test it is apparent that predictions are poor only for CSTRs 3, 4, and 5. These CSTRs each received nutrient supplements. We speculated that bacteria without the capability to decompose methyl parathion were disproportionately stimulated to grow in these systems. Their high densities led to overinflated estimates of net microbial decomposition rates. As a result EXAMS' predictions were low. We conjecture that the microbial decomposition rate coefficient for aufwuchs-dominated communities will decrease with increasing degree of eutrophication. At this point this is a qualitative conjecture because both the term "degree of eutrophication" and the function relating the rate coefficient to it have not been defined. Although needing refinement, it is a conjecture that can lead to additions to underlying theory as well as improvement in the performance of the theoretical model.

As a result of the hypothesis test, one either rejects or accepts (strictly, fails to reject) the hypothesis. As is explained in elementary statistical texts, one can make two kinds of right and two kinds of wrong decisions. One can accept a true or reject a false hypothesis, right decisions. Or one can reject a true hypothesis (type I error), or accept a false one (type II error). The probability of making a type I error is controlled by the selected rejection probability. Too often scant attention is paid to the probability of a type II error. Where the null hypothesis is equivalent to a claim that the model is correct, Balci and Sargent (1981) refer to the type I error as model builder's risk and to the type II as model user's risk. Typically, in environmental science, it is prudent to minimize model user's risk. That is, accepting a fallacious model is accompanied by too great a potential loss. A user of a faulty model could unwittingly mislead decision makers to make a decision that would result in environmental damage or significant economic loss.

Prior to the experiment, the probability of a type II error can be controlled, while holding the probability of type I error at the selected value, by planning for appropriate sample size (number of observations) and adequate replication. Failing to do so introduces the risk that a costly experiment will be run from which it is not possible to detect a faulty model. Once the experiment is complete, the probability of a type II error can only be reduced by increasing the probability of a type I error. Unfortunately many studies in environmental science suffer in the analysis stage from lack of forethought in the planning stage. For example, if an experiment is carried out without the design incorporating control over type II error, acceptance of the null hypothesis leaves unanswered a question of the likelihood that a type II error has been made. An optimal design will apportion resources to achieve a powerful test of the hypothesis. It will account for factors of the experiment that could otherwise introduce variation that would reduce the power of the test. Power is the complement to type II error; that is, it is the probability of rejecting the null hypothesis when it is false. The particular experimental

design, the underlying assumptions, and experimental error together determine the power. It is important to adhere closely to the design, so that the power in practice is equivalent to the designed power. Often as a practical matter, however, some assumptions are not met. It is useful, in these cases, to consider the effect of not adhering to the assumptions, some being more critical to the power than others. It is also enlightening to consider the ability to detect deviations of various magnitudes from the null hypothesis, using estimated error variance for the experiment and assuming that all assumptions of the design are met. In testing EXAMS we examined the power to detect deviations from the relationship of intercept = 0 and slope = 1 between predicted and observed concentrations. As an example, estimated contours of power for the test of concept validity are given in Figure 3. These contours are calculated with probability of type I error at 0.05. One can read the probability of rejecting the hypothesis when the true intercept and slope take any value. Exploration of contours of this sort can be of great value in deciding whether to reject, when the choice is not entirely clear, based upon other considerations.

SUMMARY

Testing models of the environmental fate of chemicals ultimately must provide a basis for judging their adequacy for a given purpose. One purpose for which models commonly need to be used is to aid in assessing the probable consequences of releases of new chemicals. In such cases it is evident that there are no direct measurements of their fate. Theoretical models can be used for this purpose, because they do not depend upon prior observation for their ability to predict. They depend upon the existence of relevant theory, and they must be a faithful representation of that theory. Various factors of their design, the planned uses, and the nature of the environment influence their utility. The way that these factors influence the models' adequacy can be discerned only through critical testing and interpretation. In formulating, indeed in recognizing, the tests that need to be made, it is important to comprehend how models relate to theory, and of the necessity for that relationship. It is necessary, but not sufficient, that a theoretical model be a faithful representation of theory if it is to be used to make predictions of the sort useful in assessments of chemicals. It suffices that a model be adequate for the purpose to which it is put (Overton, 1977). Adequacy can only be specified by the user, but at least the model must make use of available information to provide relevant information that does not mislead. In the final analysis the user must judge whether the model is adequate, although general criteria of usability are often discernible by the developer. Such was the question of invariance of biological decomposition rate coefficient tested for EXAMS.

Figure 4 is an illustration of a basic pattern, as I perceive it, that is followed at any level of testing. A relevant hypothesis about the model is formulated and subjected to

critical tests. Depending upon the outcome one can perceive whether the hypothesis should be rejected. If so, it can be revised and subjected to further testing or it can be abandoned. If not, it can be accepted, or it can be subjected to still further testing, possibly suggested from results of its use. In Figure 3 this pattern is repeated in the context of the three levels of testing as presented in this paper. For models that depend upon theory for their capability to predict, tests are required of theory, of the fidelity with which models represent theory, and at the utility level. For models that depend upon an empirical fit to a data set for their predictive capability, it is necessary only to test at the level of utility. It is likely that empirical models will be tested for their utility. In the environmental sciences, however, there is a continuing need to test theoretical models for their faithfulness to theory, and especially to recognize constructs that represent new hypotheses, and which, therefore, require testing at the theory level.

REFERENCES

- Balci, D. and R. G. Sargent, 1981. A methodology for cost-risk analysis in the statistical validation of simulation models. *Comm. ACM* 24(11):190-197.
- Burns, L. A., D. M. Cline, and R. R. Lassiter, 1981. User Manual for the Exposure Analysis Modeling System, EXAMS. EPA, Athens Environmental Research Laboratory, in review.
- Caswell, H., 1976. The validation problem. In: *Systems Analysis and Simulation in Ecology*, B. G. Patten (ed) Academic Press, New York, N. Y., p 313-325.
- Lassiter, R. R., R. Parrish, and L. A. Burns. Predictability of the fate of chemicals in aquatic environments. in prep.
- Paris, D. F., W. C. Steen, G. L. Baughman, and J. T. Barnett, Jr., 1981. Second order model to predict microbial degradation of organic compounds in natural waters. *Appl. Env. Microb.* 4:603-609.
- Polya, G., 1954. *Patterns of Plausible Inference*, vol II of *Mathematics and Plausible Reasoning*. Princeton University Press, Princeton, N. J. 190 p.
- Popper, K. R., 1959. *The Logic Of Scientific Discovery*. Harper and Row, New York, N. Y. 480 p.
- Popper, K. R., 1962. *Conjectures and Refutations, the Growth of Scientific Knowledge*. Harper and Row, New York, N. Y. 417 p.
- Overton, W. S., 1977. A strategy of model construction. In: *Ecosystem Modeling in Theory and Practice*, C. A. S. Hall and J. W. Day, Jr. (eds) John Wiley and Sons, New York, N. Y. p 50-73.

Table 1. Tests of intercept = 0 and slope = 1 from regression of observed on predicted concentration for validity of concept.

		parameter value	t	P > t
n=20 2 r = .87	intercept	2.75	1.18	.25
	slope	.91	1.10	.29

Table 2. Tests of intercept = 0 and slope = 1 from regression of observed on predicted concentrations for invariance of biological decomposition parameter.

		parameter value	t	P > t
n=71 2 r = .71	intercept	7.10	3.40	<.01
	slope	.83	2.67	.01

Table 3. Association of treatment factors with variation in differences between predicted and observed concentrations of MP resulting from the use of averaged biological decomposition rate coefficients to test their invariance

Source	df	F	P > F
Total	70		
Rep	1	25.53	(.01
CSTR	7	4.11	.04
Rep X CSTR	7	67.06	(.01
Date	3	27.92	.01
Rep X Date	3	115.31	(.01
CSTR X Date	21	4.25	(.01
Rep X CSTR X Date	21	27.14	(.01
Error	7		

Table 4. Duncan Multiple Range tests comparing differences (z) between predicted and observed resulting from the use of averaged biological decomposition rate coefficients

Averaged over sampling dates and arranged by CSTR

CSTR	3	4	5	2	6	7	8	1
n	10	10	9	8	8	8	8	10
mean z	9.30	5.82	3.69	-.68	-.90	-1.00	-1.28	-1.58

Figure 1. Observed versus predicted concentrations plotted with the line of no difference. Data are from the experiment to test concept validity.

Figure 2. Observed versus predicted concentrations plotted with the line of no difference. Data are from the experiment to test situation independence of the biological decomposition rate coefficient.

Figure 3. Estimated power contours for values of the slope and intercept. For null hypothesis of intercept = 0 and slope = 1, the curves represent the probability of rejecting the hypothesis should actual intercept and slope be other values. Probability of a type I error is set at 0.05. Note that at intercept = 0 and slope = 1 the probability of rejection is 0.05, the probability of type I error.

Figure 4. The general pattern of activities in testing a model. Block labeled "model" implies some testable hypothesis about a model. Block labeled "OK?" implies a decision based upon test results. Dashed lines indicate potential decision to recycle through the revision and testing without having rejected a hypothesis.

Figure 5. The general structure of model testing. Note the three levels. Note also short circuits to "accept and use" for both empirical models and theoretical models.

Figure 1

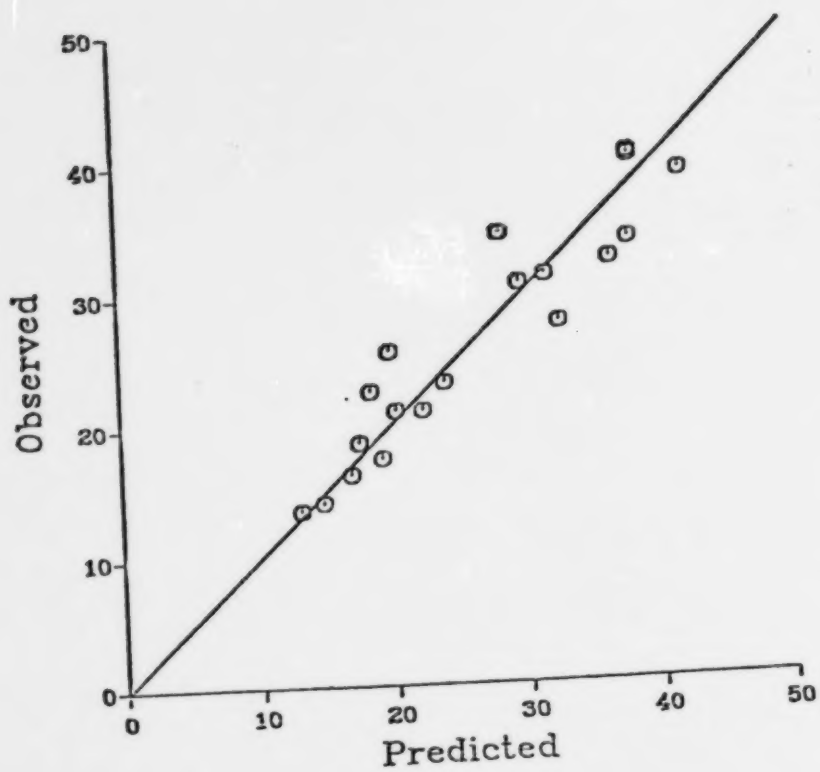


Figure 2

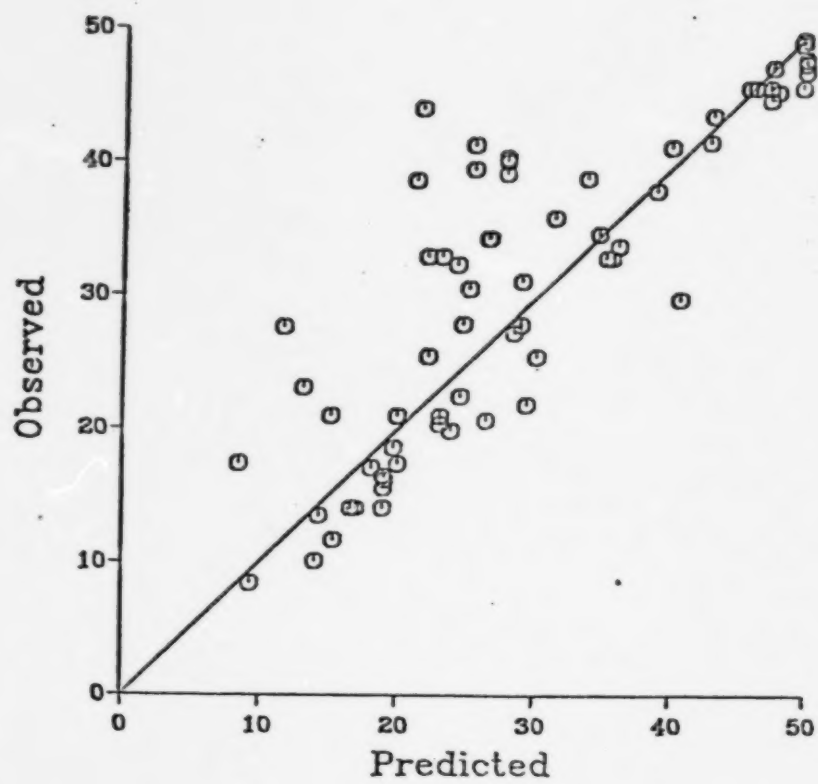


Figure 3

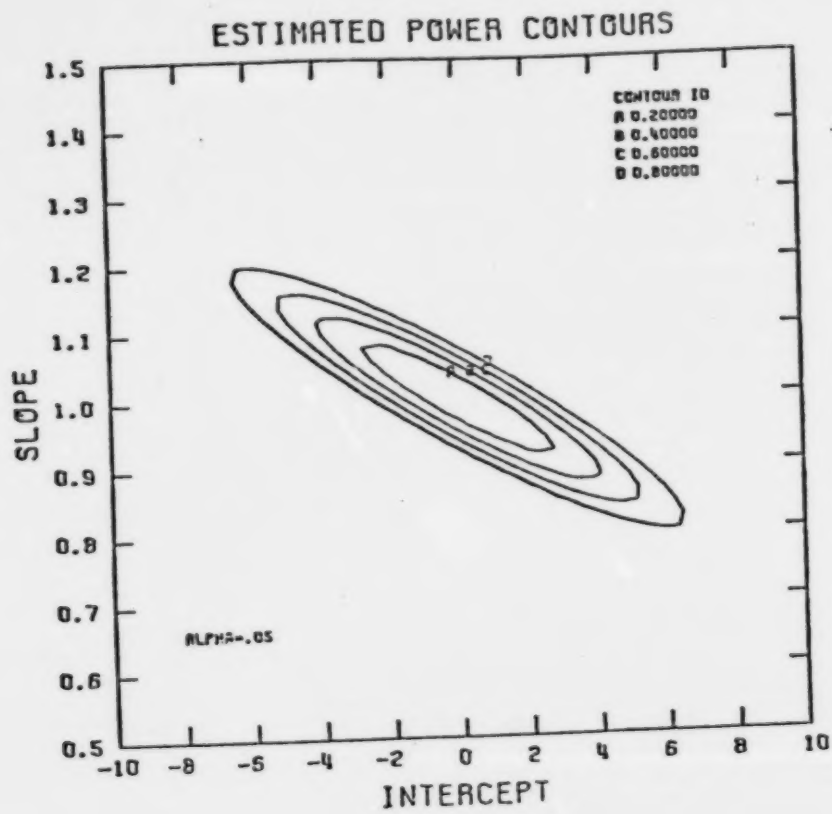


Figure 4

Test Pattern

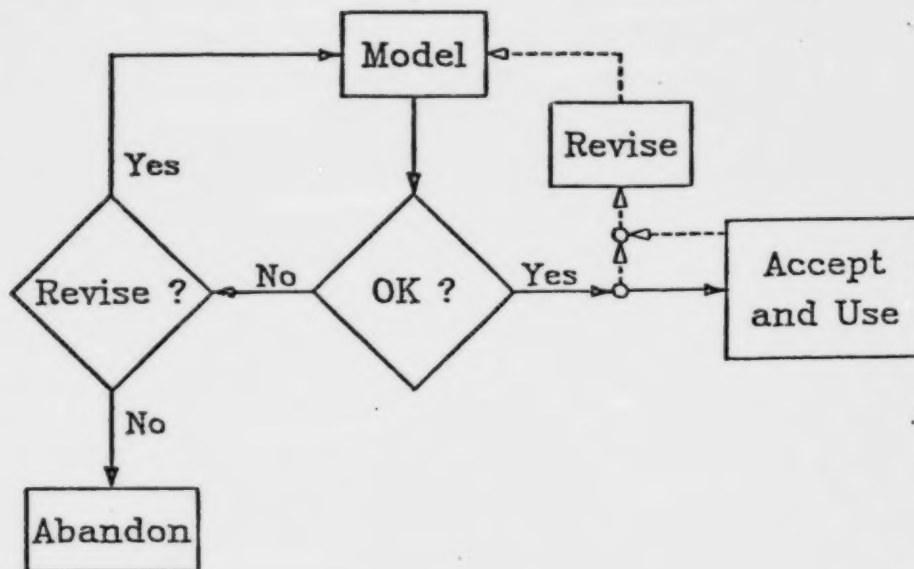


Figure 5

